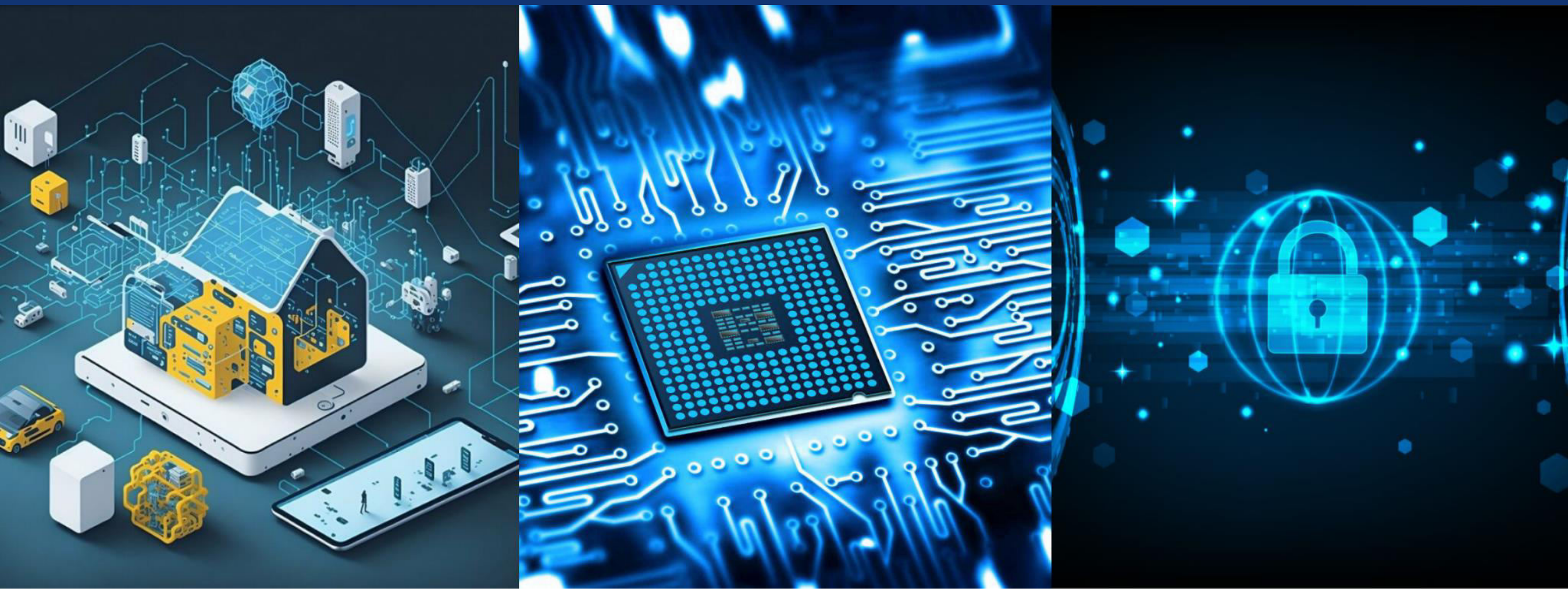
IJIRCCE

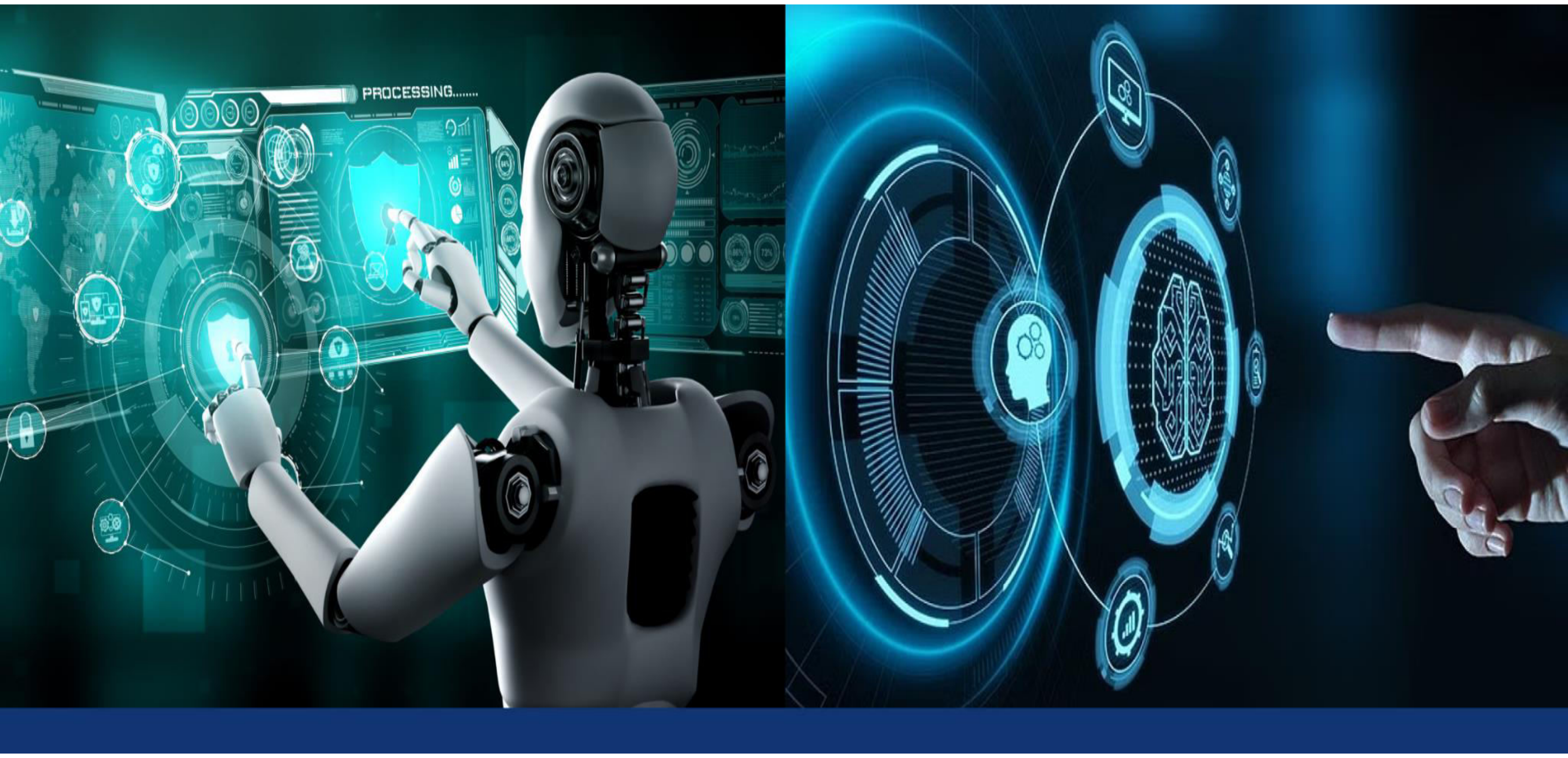
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304062

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Web-Based Smart Insurance Fraud Detection Using Machine Learning Techniques

R. Durairam[1], M. Keerthi[2], V. Alamelumangai[3], R.S. Eughendhi[4], T.K. Gowsika[5]

Assistant Professor, Dept. of CSE., Mahendra Institute of Technology (Autonomous), Namakkal, India[1]

UG Students, Dept. of CSE., Mahendra Institute of Technology (Autonomous), Namakkal, India[2,3,4,5]

ABSTRACT: Insurance Companies working as commercial undertakings for the final few a long time have been encountering extortion cases for all sorts of claims. The sum claimed by extortion is essentially gigantic and may cause genuine issues, thus along with the government, distinctive organizations working to identify and decrease such exercises. Such fakes happened in all regions of protections claims with tall seriousness such as protections claims towards the auto division is an extortion that is broadly claimed and conspicuous sort, which can be done by fake accident claim. So, we point to create a extend that works on protections claim datasets to identify extortion and fake claim sums. The project implements machine learning algorithms to make model to label and classify claim. Also, to study conducts a comparative analysis of various machine learning algorithms for classification, evaluating their performance using a confusion matrix in terms of accuracy, precision, recall, and other relevant metrics. For fraudulent transaction validation, a machine learning model is implemented using the PySpark Python library.

KEYWORDS: Machine Learning Algorithm, Web-based Application, Fraud Case location, Classifications.

I. INTRODUCTION

The act of insurance fraud involves deliberate deception aimed at unlawfully obtaining benefits from an insurance provider, posing a significant challenge to the integrity and sustainability of the insurance industry. It's a severe and critical problem that's a trouble because fraudulent insurance operations put a lesser fiscal strain on the society through high decoration prices. Recent exploration suggests that there's universal agreement that traditional styles of fraud identification are largely unreliable and squishy. These worries prompt the machine literacy and data analytics community to concentrate on this issue and seek a result. analogous to this, our proposed work directly distinguishes between fraudulent and non-fraudulent claims so that only fraudulent cases need to be delved and licit claims can be made snappily without wasting time or coffers. This study proposes an optimized and efficient approach to detecting fraudulent claims with high accuracy. A major challenge in fraud detection lies in the vast volume of insurance claims processed by companies daily. However, this challenge can be leveraged as an opportunity by integrating extensive claim databases, enabling the development of more robust and intelligent models to identify suspicious claims effectively.

We've discovered a significant issue with insurance fraud in this design. Claims filed to deceive an insurance company are known as false content claims. Since the morning of the insurance sector, there has been a patient problem with insurance fraud, with a significant portion of entered claims being fake. Insurance enterprises suffer fiscal losses from fraudulent claims, and policyholders pay advanced decorations. Machine literacy algorithms, which use data mining and deep literacy ways, help identify patterns and anomalies in insurance claim data that may be signs of fraudulent conduct. These algorithms have the eventuality to increase the delicacy significantly. These advanced styles present the insurance sector with a feasible way to ameliorate fraud discovery and lessen the goods of fraudulent claims.

Insurance companies could save plutocrat, and consumers would feel safer if these algorithms significantly bettered the delicacy and efficacity of fraud discovery. These algorithms dissect colourful aspects of claim data, similar as the kind of claim, policyholder information, and previous claim history, to spot abnormalities or questionable patterns. Insurance enterprises can produce prophetic models that use machine literacy to assign a Fraud Probability Score (FPS) to each claim. This exploration primarily focuses on using machine learning to detect fraud with auto insurances.

II. RELATED WORK

This project focuses on improving fraud detection by implementing a machine learning-based system that enhances accuracy and efficiency. Traditional fraud detection relies on manual claim verification, which is slow and prone to errors. By automating the process, the proposed system aims to detect fraud in real-time and reduce financial losses for insurers. The system preprocesses insurance claim data, extracts relevant features, and applies classification models to identify fraudulent claims. Techniques such as feature selection and resampling ensure model robustness and accuracy. Additionally, the web-based platform allows seamless fraud detection and easy accessibility for insurers. By reducing dependency on manual intervention, this system streamlines fraud detection and enhances operational efficiency. Future enhancements will involve refining fraud indicators and incorporating advanced analytics to further improve fraud detection accuracy. This research contributes to creating a reliable, automated, and scalable fraud detection framework for the insurance industry.

III. METHODOLOGY

Principally businesses ought to gain the responses to prevent fraud from passing or if that’s out of the question, to watch it before important damage is finished at intervals 407 the strategy. In utmost of the companies, fraud is understood entirely once it happens. Measures are also executed to avert it from passing over again. At intervals the given time that they can’t repel at different time intervals, but Fraud detection is that the most effective suited issue for removing it from the atmosphere and preventing from continuance. Preliminarily frauds related insurance detected or anatomized by manually using this system is not correct way to detection frauds related insurance, thus adding detection, perfection, recall we proposed this design using machine learning algorithm. Detecting fraudulent claims is crucial, as they cause major financial damage to insurance firms. To address this issue, our project implements a machine learning-based approach for accurate fraud detection.

The generic flow of machine learning data model is presented below: An intertwined exploration approach will be applied for this project. To explain the findings and conclusions of the paper and the various results, the project will employ some explicatory exploration methods in addition to experimental exploration methods, some qualitative exploration methods, andsome quantitative research methods. We’ll start by determining what is crucial for managing the data in agreement with the business that may use the model or the solutions. In this situation, an insurance provider will presumably prioritize the financial aspects of each claim while not consider particular information when developing the model. The report will detail the available data, the numerous rates, and how each traitrelates to determining whether or not this claim is fraudulent. What are the different forms of data that are available, and can the being data be bettered or changed without affecting the outgrowth of the end thing, which is relating dubious claims? To do this, data cleaning is necessary. Some values or characteristics may need to be removed, or new values may need to be created by merging existing ones and using data integration ways.

IV. XGBOOST

XGBoost is an open-source software library thatimplements optimized distributed grade boostingmachine learning algorithms under the Grade Boosting framework. XGBoost, which stands for ExtremeGradient Boosting, is widely usedfor its efficiency,scalability and ability to enhance predictive performance through gradient-boosted decision treesmachine learning library.

It provides resemblant tree boosting andis the leading machine learning library for regression classification, and ranking problems. It’s vital to an understanding of XGBoost to first grasp the machine learning generalities and algorithms that XGBoost builds upon supervised machine learning, decision trees, ensemble learning, and gradient boosting Supervised machine learning uses algorithms to train a model to find patterns in a dataset with labels and features and also uses the trained model to predict the labels on a new dataset’s features.

V. SYSTEM ARCHITECTURE

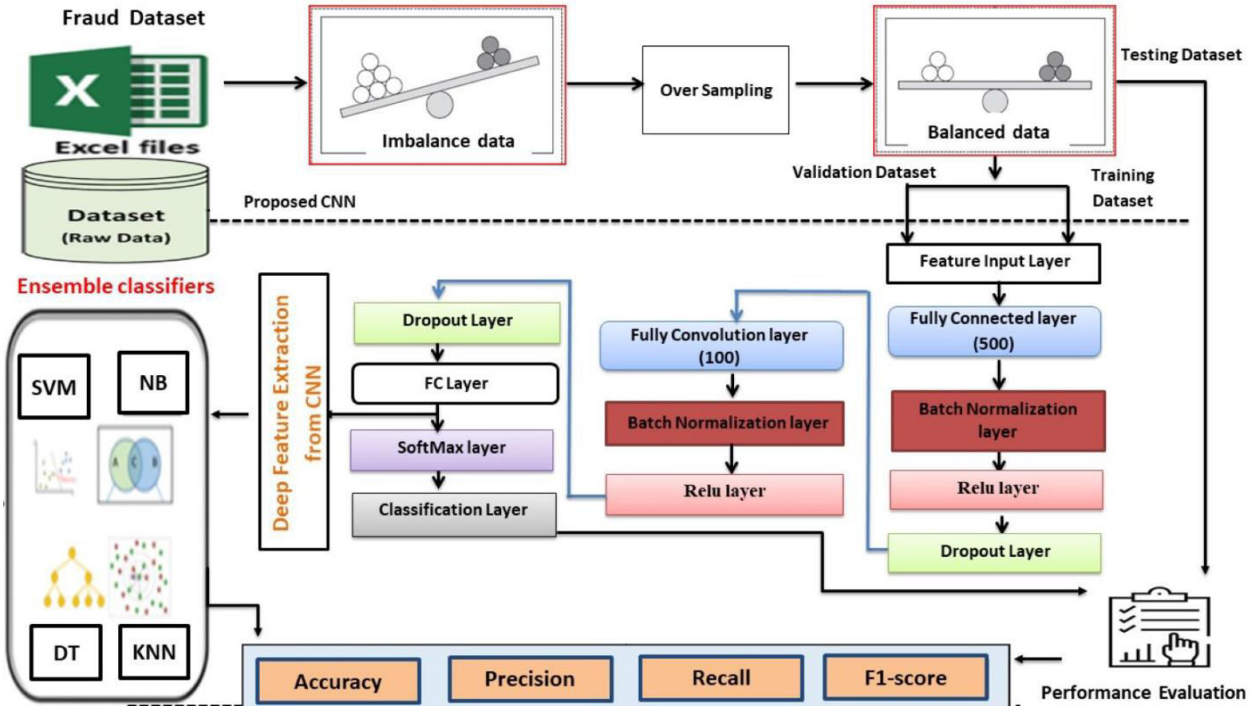
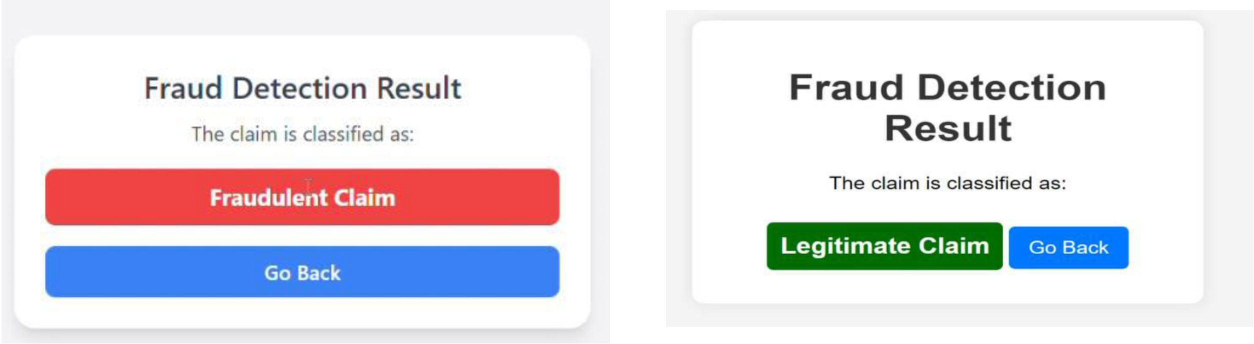


Fig. 1: Enhancing fraud detection in auto insurance and credit card transactions

One of the most widely adopted machine learning techniques has been selected for this study to ensure effective fraud detection. These algorithms produced the highest accuracy in projected results and periodic charges, amounting to billions of dollars, proving their connection to our dataset. Insurance fraud can take multitudinous forms in different insurance realms, and it can range in strictness from small-scale claim embellishment to deliberate acts of destruction or detriment. Auto insurance fraud is one of the insurers' most significant and well- known problems. A claims agent should look at costs due to fraudulent claims, which highlights the significance of securing between genuine and fraudulent claims. Although a claims agent should look into each case separately, this is constantly a precious, time-consuming, and inefficient procedure. Examining all of the numerous claims that are filed every day would be very impossible. To detect and mitigate fraudulent claims, machine learning offers a practical, quick, and provident result.

VI. RESULT ANALYSIS

A machine learning-based model was developed using various algorithms to classify insurance claims as either fraudulent or legitimate. The model was trained using historical claim data and tested on unseen records to evaluate its performance. Initially, a raw insurance claim dataset was collected and subjected to a detailed preprocessing phase. This included handling missing values, removing outliers, reducing data redundancy, and applying Min-Max normalization to ensure uniform data scaling. Categorical variables were encoded appropriately, and the dataset was divided into training and testing sets. The model training was carried out using machine learning algorithms such as Random Forest, Decision Tree, and Artificial Neural Network (ANN). Feature selection techniques were employed to identify the most relevant attributes, enhancing the model's accuracy and computational efficiency. Modelperformance was evaluated using key metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The model demonstrating the best results was selected as the optimal solution. The final system effectively detects fraudulent insurance claims in real time, contributing to enhanced accuracy and improved decision-making in claim processing.



Fraud Detection Result

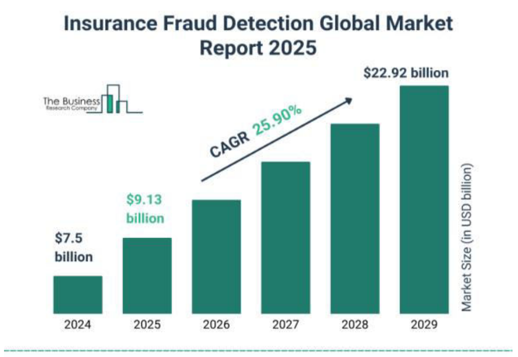
The claim is classified as:

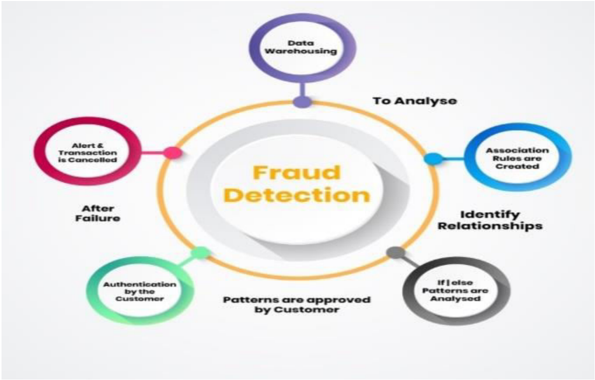
Fraud Detection Result

The claim is classified as:

Fig.2: User Interface Displaying Insurance Claim Classified as Fraudulent or Legitimate

The fraud detection system identifies potentially fraudulent insurance claims using machine learning algorithms. This a case where the system classifies a claim as fraudulent, displaying a red warning button to alert insurance professionals. This notification allows insurers to investigate the claim further before processing. This feature helps minimize fraudulent payouts and enhances fraud prevention strategies.The system also detects and verifies legitimate insurance claims efficiently, demonstrates an instance where the claim is classified as legitimate, displaying agreen confirmation button for clarity. This ensures that valid claims are processed smoothly without unnecessary delays.This streamlined approach improves accuracy, reduces manual verification efforts, and enhances trust in fraud detection automation.





Data Warehousing
To Analyse
Alert & Transaction is Cancelled
Association Rules are Created
Fraud Detection
After Failure
Identify Relationships
Authentication by the Customer
Patterns are approved by Customer
If | else Patterns are Analysed

Fig.3: Insurance Fraud Detection Market Report 2025Fig. 4: AI in Insurance Industry: Benefits & Use Cases

The machine learning models on the insurance claim dataset demonstrated effective identification of fraudulent claims, with a notably low false positive rate. This indicates satisfactory classification performance with reasonable precision. However, some datasets posed challenges due to data quality issues, which impacted prediction accuracy and model robustness. Given the intrinsic variability in different datasets, it may not be feasible to define a universally optimal algorithm or a singular feature engineering strategy to achieve consistently high performance. Instead, the choice of model and algorithm must be tailored to the specific business context and user priorities. These models are instrumental in enabling loss management units to focus on evolving fraud patterns while ensuring that the detection systems remain adaptive to new threats. Based on back-testing and the ability to identify emerging fraudulent activities, the deployed models can be considered a practical and efficient solution within the domain of insurance claim fraud detection. For future research, it is recommended to compare the performance of machine learning and deep learning techniques using more recent or advanced datasets. The current dataset used in this study is relatively imbalanced, which can affect the generalizability of the model. Therefore, utilizingbalanced and high-quality datasets is essential to improve performance evaluation. Furthermore, future work should explore feature selection techniques more comprehensively to assess the contribution of individual features to model accuracy. Identifying the variance between total features and selected subsets can help in developing more generalized and efficient models. This will ultimately enhance the reliability and scalability of fraud detection systems in real-world insurance applications.

VII. CONCLUSION AND FUTURE ENHANCEMENT

As nations increasingly shift toward economic development and digital transformation, ensuring financial integrity has become a critical priority. Detecting fraud and money laundering was once a complex task, but advancements in Machine Learning (ML) and Artificial Intelligence (AI) have significantly improved fraud detection capabilities. This study proposes a machine learning-based solution for insurance companies to efficiently detect fraudulent insurance claims. The model was developed after evaluating multiple algorithms to identify the most accurate and scalable solution. The goal was to design a model that can handle large datasets efficiently while maintaining high performance in terms of classification accuracy.A dataset containing over 1,000 insurance claim records was utilized, with a clear division into training and testing subsets. Comparative analysis showed that Random Forest and XGBoost performed better than K-Nearest Neighbours (KNN) in terms of accuracy, precision, and recall. The model can be further customized and integrated into real-time systems for automated fraud detection. This work demonstrates the effectiveness of machine learning in enhancing insurance claim validation processes and highlights its potential to significantly reduce financial losses, improve system security, and support economic growth through smarter technology-driven solutions.

Future advancements in fraud detection aim to improve accuracy, efficiency, and security. Deep learning models like CNNs and RNNs can enhance fraud pattern recognition, while blockchain integration ensures secure claim processing through transparent transaction records and smart contracts. Adaptive fraud detection using reinforcement learning enables models to evolve with emerging fraud techniques, while integration with external fraud databases improves detection accuracy by cross-referencing claim histories. Additionally, deploying fraud detection on edge devices will allow real-time processing, reducing latency and improving detection speed. These enhancements will create a more reliable, secure, and scalable fraud detection system for the insurance industry**.**

REFERENCES

1. X. Liu, J.-B. Yang, D.-L. Xu, K. Derrick, C. Stubbs, and M. Stockdale, “Automobile Insurance Fraud Detection using the Evidential Reasoning Approach and Data Driven Inferential Modelling,” 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Jul. 2020.
2. K. Ulaga Priya and S. Pushpa, “A Survey on Fraud Analytics Using Predictive Model in Insurance Claims,” Int. J. Pure Appl. Math., vol. 114, no. 7, pp. 755–767, 2017.
3. E. B. Belhadji, G. Dionne, and F. Tarkhani, “A Model for the Detection of Insurance Fraud,” Geneva Pap. Risk Insure. Issues Pract., vol. 25, no. 4, pp. 517– 538, 2000, doi: 10.1111/1468-0440.00080.
4. “Predictive Analysis for Fraud Detection.” https://www.wipro.com/analytics/comparative- analysis of-machine-learning-techniques-for-%0Adetectin/.
5. F. C. Li, P. K. Wang, and G. E. Wang, “Comparison of the primitive classifiers with extreme learning machine in credit scoring,” IEEM 2009 - IEEE Int. Conf. Ind. Eng. Eng. Manag., vol. 2, no. 4, pp. 685– 688, 2009, doi: 10.1109/IEEM.2009.5373241.
6. V. Khadse, P. N. Mahalle, and S. V. Biraris, “An Empirical Comparison of Supervised Machine Learning Algorithms for Internet of Things Data,” Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018,pp.16,2018,doi:10.1109/ICCUBEA.2018.8697476.
7. Durairam.R.. Machine Learning Approaches for Brain Disease Diagnosis. Volume 10, Issue 6, pp: 1092-1097.
8. R. Durairam Deduplication of Storage Drives Using Cloud Computing. Volume 10, Issue 6, pp: 171-172.
9. Efficient Biometric Security System Using Intra-Class Finger-Knuckle Pose Variation Assessment Mr.J.Stanly Jayaprakash, Dr.S.Arumugam, India International Journal of Computer Science & Engineering Technology (IJCSET) 2014.
10. C. Anusuya et al., "Alzheimer Disease with Blood Plasma Proteins detected using Convolutional Neural Network (CNN)," Int. J. Innov. Res. Compute. Commun. Eng., vol. 11, no. 3, Mar. 2023.
11. Parvathi M “Sensing of Near Duplicates in Large Image Database”, Volume 12, Issue 3, March 2023, DOI:10.15680/IJIRSET.2023.1203126.
12. Multimodal finger biometric score fusion verification using coarse grained distribution function JS Jayaprakash, S Arumugam2015.
13. C. Anusuya et al., "A Comparative Feature Extraction Study Using Textural Features to Extract Vital Information from Lung Images," *Int. J. Adv. Inf. Sci. Technol.*, vol. 12, no. 02, Feb. 2023.

14. Software Engineering, N. Karthigavani, Dr.K.SaravananDr.R.Vasanthi,Dr.J.StanlyJayaprakash,Dr.A.Kanchana 2018.
15. Meiyalakan K.” A Trends of Event Detection by Analyzing Social Media Platforms Data”,10.15680/ IJIRSET 2023.1203081.
16. C. Anusuya et al., "Classification of Uncertain Data Using Selection Algorithm," Int. J. Mod. Eng. Res., vol. 2, no. 3, pp. 1066-1072, May-Jun. 2012.
17. “Virtual Human Resource Management with Recruitment in Software Engineering Roles” in International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Sowmiya. R Volume 12, Issue 3, March 2024.
18. A novel approach for fingerprint sparse coding analysis using k-svd learning technique S Arthi, J Stanly Jayaprakash 2024.
19. “Intelligent Phishing Website Detection model with Deep Learning- based Innovative Technique” in International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET) Sowmiya.R, Volume13, Issue 3, March 2024.
20. Meiyalakan K “Cross-Analysis of Network Intrusion Detection Systems Based on Machine Learning”, https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook 2024.pdf.
21. “Securing personal information using data mining in Public networks”, International Journal of Computer Science and Engineering, Ms. R. Sowmiya 2017.
22. “Soil parameter analyzing using curse of dimensionality for accuracy and prediction”, in international journal of innovative research in science, engineering and technology, Ms. R. Sowmiya Feb 2020.
23. Meiyalakan K. Published following article. Online Multi-Crop Procurement and Loan System. Volume 10, Issue 5, pp: 32-35,
24. Meiyalakan K. Published following article. Knowledge-Based Approach to Detect Potentially Risky. Websites. Volume 10, Issue 6, pp: 1353-1357.
25. “Advanced Drowsiness Detection system using OPENCV and KERAS” in International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE) Sowmiya.R, Volume 12, Issue 5, May2024.
26. Meiyalakan K.” Estimation of the Available Bandwidth in Inter-Cloud Links for Task Scheduling in Hybrid Clouds”, DOI: 10.15680/IJIRCCE.2022.1007086
27. Real time cyber physical false data attack Detection in machine learning methods in International Journal of Innovative Research in Computer and Communication Engineering, Sowmiya.R, June 2021.
28. C. Anusuya et al., "Diagnostics Decision Support System for Tuberculosis Using Fuzzy Logic," IRACST Int.J. Compute. Sci. Inf. Technol. Sec., vol. 2, no. 3, Jun. 2012.
29. “Predicting Emotions from Text Using Computing Technique” in the “International conference on Integrating Recent Innovations in Science and Technology: Shaping the future (ICIRIST-2024)” Sowmiya.R,2024.
30. Meiyalakan K. “Machine Learning Contributions to the Field of Security and Privacy Using Android”, ICSIEM 2024,16 -17, April 2024.
31. *C.* Anusuyaet al "Credit Card Fraud Detection using Machine Learning Based Random Forest Algorithm, "Int. Sci. Adv.Res. Technol., vol. 9, no. Mar.2023.
32. Deep Q-Network with Reinforcement Learning for Fault Detection in Cyber-Physical Systems J. Stanly JayaprakashM. Jasmine Pemeena Priyadarsini, B.D. Parameshachari Hamid Reza Karimi, and Sasikumar GurumoorthyJournal of Circuits, Systems and Computers 2022.
33. [34] C. Anusuya et al., "Facial Recognition Services for E-Voting System by Using Blockchain Technology," Int. J. Innov. Res. Compute. Commun. Eng., vol. xx, no. xx, pp. xxx-xxx, year.
34. Meiyalakan K. “Blockchain-Based Anonymous Authentication of Cloud Data Using Stochastic Diffusion SearchAlgorithm”,https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf.
35. Energy and Green IT Resource Management Analysis and Formation in Geographically Distributed Environmental Cloud Data Centre” Murugan G, Gayathri.C, Latha.S, Sathiya Kumar C, SudhakarSengan, PriyaV(2020),in International Journal of Advanced Science and Technology Vol. 29, No. 6,pp 4144-4155(SCOPUS indexed).
36. Cloud Data Encryption and Authentication Based on Enhanced MerkleHash TreeMethod J.Stanly Jayaprakash 1, Kishore Balasubramanian2, Rossilawati Sulaiman 3,MohammadKamrulHasan3,*,B.DParameshachari4 and Celestine Iwendi 2021.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 ✆ 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com

Scan to save the contact details